



Linguistique et corpus spécialisé : l'exemple de l'alternance nominalisation/verbe

Anne Condamines

► To cite this version:

Anne Condamines. Linguistique et corpus spécialisé : l'exemple de l'alternance nominalisation/verbe. Journées Scientifiques du CIRLEP (Centre Interdisciplinaire de Recherche sur les Langues Et la Pensée), 2001, Reims, France. pp.237-258. halshs-01380888

HAL Id: halshs-01380888

<https://shs.hal.science/halshs-01380888>

Submitted on 17 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linguistique et corpus spécialisé : l'exemple de l'alternance nominalisation/verbe

Anne condamines
ERSS, UMR 5610 CNRS
5 allées Antonio Machado
31058 Toulouse cedex
tél : 05 61 50 36 08
mél : acondam@univ-tlse2.fr

1. Introduction

L'un des aspects de l'opposition verbo-nominale qui a été le plus étudié dans les travaux de linguistique est sans doute l'opposition nominalisation déverbale/verbe. Que ce soit en syntaxe, en sémantique ou en morphologie, nombre de travaux évoquent cette opposition souvent d'ailleurs en la considérant comme une alternance plutôt que comme une réelle opposition. La question qui se pose au fond est celle de savoir si ces deux catégories grammaticales, formellement apparentées, correspondent ou non à la même signification. Les études que nous présentons ici ont ceci de particulier qu'elles sont réalisées sur des corpus et plus précisément sur des corpus spécialisés ; elles ne se situent donc pas dans une visée de description d'un système linguistique sur la base d'une réflexion introspective mais prennent appui sur des données textuelles attestées. L'article présente ainsi à la fois des résultats d'analyse sur des corpus spécialisés et des questions nouvelles que peuvent amener ces études du fait même qu'elles sont réalisées sur des données attestées, plutôt que sur des données introspectives. En effet, l'utilisation des corpus en linguistique est sans doute en train de renouveler les méthodes de cette discipline et ce renouvellement est en partie amené par les travaux de chercheurs travaillant sur les corpus spécialisés qui ne peuvent faire seulement appel à leur intuition linguistique. Nous évoquerons dans une première partie d'une part la problématique de l'analyse de corpus et d'autre part celle des corpus spécialisés. Les parties suivantes rendront compte de résultats obtenus lors de différentes études : sur les nominalisations et leur dénombrement en corpus spécialisé, sur les nominalisations dans différentes structures syntagmatiques, enfin, sur l'alternance nominalisation/verbe proprement dite.

2. Problématique de l'analyse de corpus spécialisés

L'analyse de corpus, qui sous-entend une analyse d'usage réels, a été longtemps ignorée, voire méprisée par une grande partie des chercheurs en linguistique, largement influencés par l'approche générativiste. Seule peut-être, la sociolinguistique a utilisé de façon systématique les corpus, le plus souvent oraux, comme matériau de travail. L'essor actuel des travaux de linguistique à partir de corpus est dû au développement des technologies qui permet de mettre à disposition d'une part de grands volumes de données textuelles et d'autre part, des outils pour archiver, interroger et analyser ces données. La linguistique se trouve ainsi obligée d'intégrer dans sa problématique les approches sur corpus, (au lieu de les cantonner aux marges des sciences du langage, du côté de la linguistique « appliquée »), ce qui ne va pas sans lui poser des questions fondamentales.

Lorsque les corpus sont « spécialisés », l'analyse se confronte à celles d'autres disciplines, comme la terminologie et le linguiste doit défendre une approche qui prend appui sur l'analyse de corpus plutôt que sur la dénomination de référents extra-linguistique. Le fait que ces corpus soient reliés à un domaine particulier constitue une des caractéristiques qui sont prises en compte dans l'analyse, au même titre que le genre textuel du corpus, le degré d'expertise du locuteur et de l'interlocuteur...

2.1. Les corpus spécialisés

La tradition terminologique, dans la continuité des travaux de E. Wuster (Wuster, 1981), considère que la langue scientifique et technique a un fonctionnement particulier qui la met à l'abri des imperfections de la langue générale que sont la synonymie, la polysémie et l'ambiguïté. Malheureusement, cette vision idéalisée ne résiste pas à l'analyse et de nombreux travaux tendent à montrer que les corpus spécialisés mettent en œuvre des mécanismes tout à fait comparables à ceux d'autres types de corpus.

Toujours dans cette vision traditionnelle des corpus spécialisés, seuls les noms sont pris en considération car ils constituent l'élément le mieux adapté pour dénommer des concepts, censés exister préalablement. Là encore, l'étude approfondie de corpus spécialisés met en évidence des fonctionnements sémantico-syntaxiques qui, en réalité, concernent tous les éléments grammaticaux. En effet, dans une étude sémantique de corpus spécialisés, on ne peut confiner l'analyse aux seuls éléments lexicaux, pris isolément ; on s'intéresse obligatoirement au contexte dans lequel apparaissent ces éléments, c'est-à-dire qu'on s'intéresse au fonctionnement syntagmatique et syntaxique, ce qui conduit à une description fine de tout le corpus. François Rastier (Rastier, 1995) a montré qu'élaborer une terminologie à partir de corpus revenait à construire des types, à partir d'un ensemble d'occurrences, construction qui se fait en fonction d'un objectif particulier et à partir de l'étude des contextes de ces occurrences.

Influencés dans leur constitution initiale par la tradition terminologique, la plupart des outils d'extraction de termes-candidats ne s'intéressent qu'aux syntagmes nominaux ce qui pose un problème majeur si l'on s'intéresse aux actions, aux processus dans un corpus. Une façon de contourner ce problème, tout en bénéficiant de l'assistance de ce type d'outils, consiste à étudier les nominalisations et à partir d'elles leur alternance avec des verbes dans les corpus. Ce sont les résultats des travaux dans ce domaine menés à l'ERSS (Equipe de Recherche en Syntaxe et Sémantique) (cf les travaux de D.Bourigault, C.Fabre et A.Condamines) qui sont présentés dans les prochains paragraphes.

2.2. La linguistique de corpus

La possibilité de travailler avec des corpus en linguistique peut s'envisager de différentes façons. Dans une approche hypothético-déductive, on utilise des données attestées pour vérifier une intuition, il ne s'agit pas alors d'analyser un corpus dans un objectif précis mais de rechercher dans du texte, considéré en vrac, confirmation de son intuition. Dans une approche inductive, le corpus constitue le point de départ d'une étude : on recherche des régularités de fonctionnement afin de constituer un système. Plusieurs questions se posent : les régularités observées sont-elles propres à ce corpus ou peuvent-elles être extrapolées à l'ensemble des corpus du même type ? Si cette extrapolation est possible, comment caractériser le type du corpus étudié ? Soit l'étude doit ne concerner que ce corpus, par exemple dans la perspective d'étudier les terminologies utilisées par différents partenaires d'un même projet. Dans ce cas, on doit s'assurer que le corpus est « équilibré », que tous les partenaires sont également représentés, que les textes relèvent bien du même genre discursif. Soit l'étude porte sur un corpus en tant qu'il est représentatif de corpus ayant les mêmes caractéristiques et il doit être construit en conséquence : par exemple, il doit couvrir un même domaine de connaissances représenté par plusieurs genres discursifs. Se pose alors la question du rapport entre le système élaboré pour ce corpus et un système plus général que pourrait être le système de la langue, en fait, se pose le problème de la généralisation des résultats. Entre une conception qui fait d'un corpus une actualisation du système linguistique (c'est-à-dire qui neutralise les variations d'un corpus à l'autre) et une autre qui considère le corpus comme la seule unité pertinente, (c'est-à-dire qui considère qu'il n'y a pas de langue générale), beaucoup d'options sont possibles et le panorama est difficile à dresser, en particulier parce que la sémantique de corpus est une discipline jeune. Les prochaines années qui vont

certainement voir se développer les travaux dans ce domaine devraient permettre d'obtenir des résultats qui feront avancer cette problématique théorique du rôle des corpus pour la linguistique.

Quoi qu'il en soit, l'analyse de corpus suppose nécessairement la mise en œuvre d'outils et de fait, dans les dernières années, de nombreux logiciels ont été conçus pour assister l'exploration en corpus, pour la plupart des prototypes de laboratoire.

Notons aussi que dans le cas où l'analyse vise le repérage de fonctionnements qui puissent être extrapolés à d'autres corpus présentant les mêmes caractéristiques, le recours à la comparaison est inévitable. En effet, la description ne prend sa pertinence que par rapport à d'autres éléments dont les caractéristiques sont à la fois proches et différentes. Ainsi, dans les travaux présentés ici, la comparaison se fait soit entre corpus, soit, à l'intérieur d'un corpus, entre éléments grammaticaux, pour nous entre les nominalisations et les verbes. Les résultats présentés ici sont obtenus pour quelques corpus techniques mais nous envisageons de les évaluer sur des corpus différents mais ayant les mêmes caractéristiques de genre

3. Analyse du fonctionnement des nominalisations

Les deux premières études s'intéressent aux seules nominalisations, la première d'un strict point de vue statistique, la seconde du point de vue des contextes syntagmatiques dans lesquels elles apparaissent.

3.1. Approche statistique du fonctionnement des nominalisations

Cette étude est faite uniquement à partir d'une approche statistique.

3.1.1. Mise en place de l'étude

Précisons que cette étude est décrite de manière exhaustive dans (Condamines, 1998). Nous n'en donnons ici que les principaux éléments.

Le travail a été réalisé sur un corpus appelé Mougli, fourni par EDF : il s'agit d'un Manuel de consignes concernant les spécifications en génie logiciel et la rédaction de la documentation en lien avec ces spécifications. Le corpus comporte environ 50000 mots.

La première tâche a consisté à repérer les nominalisations dans le corpus. Elle a été faite avec le logiciel Nomino, qui nous a fourni la liste des noms à partir de laquelle nous avons sélectionné les nominalisations déverbiales, sur des bases intuitives (parenté morphologique avec un verbe et sens « prédicatif », par exemple *gestion/gérer*). Le corpus compte 1151 formes nominales parmi lesquelles 292 sont des nominalisations. Sur ces 292, nous avons examiné les 185 qui apparaissent plus de trois fois dans le corpus.

Trois études ont été menées. La première portait sur le dénombrement de nominalisations, grâce à l'outil Hyperbase (Brunet, 1995) qui permet de comparer le corpus d'étude avec une partie du corpus Frantext. Cet outil donne quatre types de résultats concernant les lexèmes (pour nous, les nominalisations) :

- les lexèmes hors modèle (non compris dans la liste des 10000 formes les plus fréquentes),
- les lexèmes en excédent dans le corpus d'étude,
- les lexèmes conformes au modèle,
- les lexèmes en déficit.

La deuxième étude visait le dénombrement de nominalisations par rapport aux formes verbales. Le corpus de comparaison a été, cette fois, une partie du journal le Monde et la méthode utilisée celle du calcul du χ^2 à quatre chiffres (formes nominales dans Mougli, formes nominales dans Le Monde,

formes verbales dans Mougli, formes verbales dans Le Monde). Le logiciel d'exploration de textes était Sato (un outil d'analyse de corpus construit par le Centre ATO de Montréal).

La troisième étude a, elle aussi, utilisé le corpus du Monde comme référence, le calcul du χ^2 comme méthode d'analyse et Sato comme outil d'analyse. Il s'agissait ici de comparer les formes nominales au pluriel par rapport aux formes nominales au singulier et ce, dans le corpus Mougli, en comparaison avec le corpus du Monde. L'hypothèse était qu'un nombre de formes au pluriel plus élevé dans Mougli que dans Le Monde pourrait venir d'un usage des nominalisations plus particulièrement lié à la dénomination de résultats d'actions que de processus. On sait en effet que les nominalisations utilisées pour renvoyer à un processus plutôt qu'à son résultat peuvent difficilement prendre le pluriel, comme le montre le test avec la préposition *pendant* qui force l'interprétation processive :

- * pendant les constructions des immeubles
- pendant la construction des immeubles.

3.1.2. Résultats

Les résultats obtenus peuvent être résumés de la façon suivante.

Pour l'étude du dénombrement avec Hyperbase, on trouve :

- 80 nominalisations hors modèle (hors des 10000 formes les plus fréquentes),
- 74 nominalisations en excédent,
- 31 conformes au modèle (i.e au fonctionnement dans frantext)
- 0 en déficit.

Ces chiffres sont déjà particulièrement parlants et montrent que les nominalisations sont nettement plus utilisées dans Mougli que dans Frantext ; en effet, 83 % (80 + 74) des nominalisations sont plus utilisées dans Mougli que dans Frantext, seulement 17 % (31) sont utilisées de façon similaire dans les deux corpus et aucune nominalisation n'est moins utilisée dans Mougli. Ce taux élevé de nominalisations en corpus spécialisé sera à nouveau vérifié dans les résultats présentés en 3.

Les résultats de la deuxième étude, sur le dénombrement de nominalisations par rapport aux formes verbales, semble cohérent avec les résultats précédents. En effet, lorsque la nominalisation appartient à un des deux premiers groupes (« hors modèle » ou « en excédent ») qui correspondent aux cas où les nominalisations sont nettement plus nombreuses dans le corpus spécialisé, alors cette forme nominale est nettement plus utilisée que la forme verbale. Tout se passe comme si la nominalisation était préférée à la forme verbale. Pour expliquer cet usage élevé de la forme nominale, on peut tenter deux explications :

- l'utilisation de la forme nominale permet de ne pas personnaliser le prédicat, il n'y a pas nécessité de donner un sujet ; cela peut être une façon d'atténuer la dimension injonctive que l'on trouve dans un corpus de type « manuel », particulièrement celui-ci où les énoncés ne sont pas des conseils à suivre mais quasiment des obligations ;
- comme cela a pu être souligné, la forme nominale est perçue comme la plus élaborée, censée renvoyer à un concept particulièrement stabilisé.

Les résultats de la troisième étude, sur la comparaison des nominalisations au singulier et au pluriel, montrent que seulement un quart des nominalisations a un fonctionnement différent dans Mougli et dans Le Monde, c'est-à-dire que pour 1/4 des nominalisations, l'une des formes (le singulier ou le pluriel) est plus ou moins utilisée dans Mougli que dans frantext. En tout cas, il n'y a pas un usage massif de la forme plurielle dans le corpus spécialisé. Ce résultat invalide donc en grande partie l'hypothèse que les formes nominales soient utilisées pour des artefacts principalement.

Pour synthétiser, on peut retenir de cette étude, d'une part que les nominalisations sont plus utilisées dans le corpus spécialisé que dans le corpus Frantext, d'autre part que cette sur-utilisation semble se faire au détriment des formes verbales.

3.2. Analyse du contexte syntagmatique des nominalisations

Pour cette étude (dont les résultats sont présentés dans (Condamines, 1999), le même corpus, Mougliis, a été utilisé. En revanche, le corpus de comparaison était une partie du corpus Le Monde. Trois structures, souvent considérées comme concurrentes, ont été examinées :

- Nominalisation + de + SN (*gestion de logiciel*),
- Nominalisation + N (*conception composant*),
- Nominalisation + adjectif relationnel (*configuration logicielle*).

Sur les 185 nominalisations étudiées dans la première étude, seules les 97 qui occurrent au moins 3 fois, en cumulant les trois types de structures, ont été analysées.

Le repérage de la nature syntaxique des compléments a été parfois difficile :

- distinction nom/adjectif : des exemples comportant des éléments tronqués peuvent créer une ambiguïté : *production télé* peut correspondre soit à *production de la télé* (1ère structure), soit à *production télévisuelle* (3è structure),

- distinction adjectif relationnel/adjectif qualificatif ; j'ai retenu comme critères pour les adjectifs relationnels, le fait qu'ils soient dérivés d'une base nominale et le fait qu'ils ne varient pas en degré. Ce deuxième critère est parfois difficile à contrôler. Ainsi si *fabrication industrielle* semble accepter difficilement la variation en degré (**fabrication très industrielle*), *fabrication manuelle* semble beaucoup moins réticent pour accepter cette variation : (*?*fabrication très manuelle*). Il est bien évident aussi que cette intuition peut ne pas être valide dans des corpus spécialisés qui justement, ont pour caractéristique de fonctionner différemment des usages habituels et souvent de manière inattendue.

La nature sémantique des arguments de la nominalisation a été également examinée. Plusieurs nominalisations ont comme complément une autre nominalisation, ce phénomène est très peu mentionné dans les travaux existants sur le domaine ; la nature sémantique de ces compléments-nominalisations est très difficile à évaluer : *action d'accompagnement*, *réunion d'avancement*...

Les résultats chiffrés de cette étude sont donnés dans le tableau ci-dessous :

	MOUGLIS	LE MONDE
1- Chaînes de caractères	50004	1720086
2- Nominalisations	4399 (8 % de 1)	9670 (0,5 % de 1)
3- + de N	1572 (35,5 % de 2)	2869 (30 % de 2)
4- + Adj relationnel	215 (5 % de 2)	1035 (10,5 % de 2)
5- + N	330 (7,5 % de 2)	19 (0,1 % de 2)

Commentaires.

Les adjectifs relationnels sont peu utilisés dans Mougliis comparativement au Monde. Sur 97 nominalisations, 67 ne sont jamais utilisées avec un adjectif relationnel alors que c'est le cas de seulement 33 dans Le Monde. Cette remarque peut être interprétée aussi en attribuant le rôle de corpus témoin au corpus Mougliis ce qui revient à constater une sur-utilisation de l'adjectif relationnel dans Le Monde ce qui est peut-être une caractéristique d'un corpus journalistique.

En revanche, la structure Nominalisation + N est, elle, largement plus utilisée dans Mougli. Dans la très grande majorité des cas, ce second N joue un rôle argumental d'objet pour la nominalisation (*contrôle qualité, livraison produit*). Le seul cas d'argument non-objet concerne l'exemple : *acceptation client*. Ces cas semblent correspondre à des ellipses de la préposition, les cas d'alternance de formes avec et sans préposition ne sont d'ailleurs pas rares : *documentation du produit/documentation produit, intégration du produit/intégration produit*. Dans Le Monde, l'alternance se fait, elle, entre adjectif relationnel et complément de N : *action gouvernementale/action du gouvernement, construction européenne/construction de l'Europe, modification législative/modification de la loi*. Autant d'éléments qui semblent permettre de dire que là où un corpus journalistique comme Le Monde utilise l'adjectif relationnel, un corpus spécialisé comme Mougli préfère, lui, utiliser l'ellipse de la préposition. Pour que cette hypothèse soit généralisée à un ensemble de corpus, il est bien évident qu'il serait nécessaire de refaire l'analyse sur plusieurs autres corpus ayant le même type de caractéristique.

Les quelques cas de construction directe Nominalisation Nom, dans Le Monde, correspondent à des juxtapositions de noms sans relation très nette entre eux (cf (Noailly, 1990) : *organisation mère, opération coup de poing*).

Du point de vue de la nature des arguments, certaines nominalisations (13 exactement) ont un fonctionnement très différent dans Mougli par rapport au Monde. Voici quelques exemples dans le tableau suivant.

	Complément dans MOUGLIS	Complément dans LE MONDE
Action	1- + Nominalisation (16)	1- + Agent (53) 2- + Nominalisation (21)
Analyse	1- + Objet (15)	1- + Objet (39) 2- + Agent (19)
Contrôle	1- + Objet (46)	1- + Objet (91) 2- + Agent (28)
Demande	1- + Objet (14) 2- + Agent (1)	1- + Objet (97) 2- + Agent (71)
Démarche	1- + Nominalisation (17) 2- + Agent (1)	1- + Agent (17) 2- + nominalisation (8)
Gestion	1- + Objet (297)	1- + objet (100) 2- + Agent (9)
Réunion	1- + Nominalisation (13)	1- + Agent (67) 2- + Nominalisation (13) 3- + Objet (3)

Selon les cas, un ou deux types d'arguments disparaît, la plupart du temps, l'argument *agent* (comme pour *action, analyse, contrôle, gestion, réunion*), ou bien la proportion, en fonction du type d'argument est inversée (*démarche*), ou bien, un des arguments est nettement moins utilisé (comme l'argument *agent* pour *demande*).

Quelques exemples :

Action est suivi dans Mougli des nominalisations suivantes : *de contrôle qualité, d'assurance qualité, de gestion de configuration, de regroupement de fichiers, de contrôle, de suivi, d'assistance et de formation, d'accompagnement ou d'intégration de personnes*. Il n'apparaît jamais avec un nom ayant le rôle d'agent.

Dans Le Monde, *action* est aussi suivi de nominalisations comme : *d'enseignement, de sensibilisation à la danse, de sponsoring, de prévention, d'évaluation, d'animation sportive, ...* mais aussi de noms

qui ne sont pas des nominalisations comme : *du Plan, de la justice, de l'administration, de commandos, de la France à l'étranger, des élus, de l'avocat dakarois, du premier Ministre...*

De manière générale, il est très net que l'argument *agent* a tendance à disparaître. Cette constatation corrèle une hypothèse faite dans le paragraphe précédent, celle qui concerne le fait que la plus grande utilisation de formes nominales pourrait être due à la possibilité de ne pas personnaliser le procès. La rareté de l'expression de l'agent pourrait être en lien avec le genre du document (un manuel) qui n'a pas pour but de personnaliser une action mais d'indiquer une marche à suivre ; par ailleurs, comme nous l'avons déjà dit, ce manuel a un style particulièrement injonctif.

Pour résumé, par rapport au corpus Le Monde, le corpus Mougliis utilise peu l'adjectif relationnel, qu'il remplace par une ellipse de la préposition dans la structure Nominalisation de N. Dans certains cas, la structure argumentale est très différente dans Mougliis.

4. Etude de l'alternance Nominalisation/verbe dans trois corpus spécialisés

Cette étude a été réalisée avec Didier Bourigault ; elle est présentée en détail dans (Bourigault et Condamines, 1999).

4.1. Mise en place de l'étude

Ce travail avait une visée plus sémantique que les précédents. Par ailleurs, il s'intéresse non plus seulement aux noms mais aussi aux verbes et à l'alternance de ces formes dans des corpus.

Les trois corpus qui ont servi de base à l'étude sont :

- Mougliis (50000 mots),
- GDP, un autre corpus d'EDF (220000 mots),
- Un corpus fourni par la DDE de la Haute-Garonne (147000 mots).

Les trois corpus ont été étiquetés grammaticalement par l'outil Sylex. Comme ressource de base, nous avons mise en œuvre Verbaction, qui contient environ 7000 couples nominalisation/verbe (démarrage/démarrer, démenti/démentir) et qui a été réalisé par Nabil Hathout.

Comparaison du nombre de formes nominales avec le nombre de formes verbales

Un premier examen nous a permis de comparer le fonctionnement de ces trois corpus techniques à trois corpus littéraires : un livre de Sartre, un livre de Balzac, un livre de Chateaubriand, du point de vue de la répartition nominalisations/verbes.

Rappelons que dans une étude précédente, nous avons examiné cette même répartition en comparant Mougliis avec le fonctionnement du journal Le Monde, en calculant le χ^2 . L'étude ici se fait sur le seul dénombrement d'occurrences. Les résultats présentés dans le tableau suivant sont particulièrement parlants.

	Nominalisations	Autres noms	Verbes
Sartre	1648 (8%)	9548 (48 %)	8680 (44%)
Balzac	4751 (7 %)	35070 (53 %)	26359 (40%)
Chateaubriand	941 (6%)	7808 (50 %)	6713 (44%)
DDE	12243 (25 %)	25455 (54 %)	10271 (22 %)
MOUGLIIS	5212 (20 %)	16082 (63 %)	4294 (17 %)
GDP	14293 (24 %)	32685 (55 %)	12662 (21 %)

Il s'avère qu'avec une étonnante régularité, les nominalisations sont plus utilisées dans les corpus techniques que dans les corpus littéraires (entre 20 et 25 % dans les corpus techniques, autour de 7 % dans les corpus littéraires). En revanche, les verbes sont plus utilisés dans les corpus littéraires (entre

40 et 44 %) que dans les corpus techniques (aux alentours de 20 %). La proportion d'autres noms est assez comparable dans tous les corpus. Nous aboutissons ainsi au même type de résultats que dans l'étude précédente : les nominalisations sont bien plus utilisées dans un corpus technique comme Mougli (dans tous les corpus techniques ?) que dans d'autres types de corpus.

4.2. Alternance nominalisation/verbe

Notre principal objectif pour cette étude a été d'examiner l'alternance nom/verbe directement dans les corpus où ils apparaissent. L'objectif était d'évaluer si cette alternance correspondait à des significations identiques ou non. Une interface réalisée en ACCESS nous permettait d'accéder en parallèle aux contextes des nominalisations et des verbes leur correspondant. Nous nous sommes basés sur l'examen et le classement des contextes pour décider si nous avions similitudes de fonctionnement ou pas entre les nominalisations et les verbes, i.e, si les arguments des unes et ceux des autres pouvaient être considérés comme équivalents, d'un point de vue sémantique. Ce classement des arguments pose évidemment des problèmes. La rigueur de l'analyse distributionnelle voudrait qu'il soit fait sur la base de cooccurrences comportant des éléments identiques. Les corpus n'étant pas suffisamment volumineux, cette exigence est difficile à respecter. Nous l'avons remplacé par la notion de classe sémantique, qui pose le problème d'une interprétation du linguiste et sa connaissance de la langue, connaissance qui peut ne pas être pertinente ou suffisante sur des corpus techniques.

Plusieurs cas de figures apparaissent, qui sont détaillés dans les prochains paragraphes.

4.2.1. Nom et verbe fonctionnent de façon identique

Ils acceptent le même type d'argument, on considère qu'ils ont la même signification. C'est le cas de la paire *recenser/recensement* dans le corpus DDE (respectivement 15 et 10 occurrences) où les deux membres acceptent des arguments que l'on peut classer sous le thème : « situations à gérer » : *contraintes, perturbations...*



Figure 1 : le cas de *recenser/recensement* (corpus DDE)

D'autres perturbations (météorologiques, hydrologiques) sont susceptibles d'affecter le réseau routier mais leur recensement n'a pas été produit de manière exhaustive.
Les bouchons sur le réseau ASF sont recensés par des véhicules de balisage placés en queue de bouchon.

4.2.2. Nom et verbe fonctionnent de façon complètement différente

La paire *conduite/conduire* fonctionne de cette façon dans le corpus DDE. Les arguments pour la nominalisation (66 occurrences) et pour le verbe (28 occurrences) sont très différents ; la nature même de la construction argumentale est différente : avec la préposition *de* pour le nom, avec la préposition *à* pour le verbe. On trouve ainsi des arguments comme *étude, opération, réflexion, projet* pour le nom et des arguments verbaux ou nominaux (nominalisations) comme *mise en place, rédaction du document, prévoir une implantation...* pour le verbe.

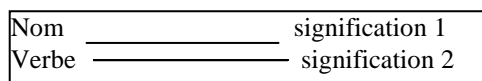


Figure 2 : le cas de (*conduire ; conduite*) dans le corpus DDE

Les personnels affectés à l'équipe de conduite d'opérations seront identifiés
L'examen des fiches de prévision peut conduire la CDES à déclarer non courant un chantier initialement courant.

4.2.3. Seule, la forme verbale est attestée

Dans GDP par exemple, seul *assurer* est attestée. Les arguments avec lesquels il est construit peuvent s'organiser en deux groupes : l'un pouvant être associé au sens *fonction*, l'autre au sens *condition optimale*. On considère donc que *assurer* a deux significations.

Le réseau HTA assure la liaison entre les jeux de barres HTA.

assurer la stabilité des groupes de production, (...) assurer une qualité de desserte satisfaisante, (...) assurer la sécurité des matériels et la meilleure exploitation possible du système production-transport

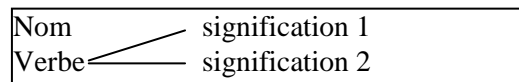


Figure 3 : le cas de (*assurance* ; *assurer*) dans le corpus GDP

4.2.4. Seule, la forme nominale est attestée

Il en va ainsi pour *perte* dans GDP. Comme dans le cas précédent, deux groupes d'arguments peuvent être distingués : l'un sous le terme générique *ouvrage*, l'autre sous celui de *énergie*. Dans ce second cas, l'argument est très souvent éliminé.

Pour réalimenter les auxiliaires de sûreté après la perte du Transformateur de Soutirage (TS), (...).

Mais une mise a la terre des écrans des deux extrémités d'un câble conduit à augmenter les pertes par courants de circulation de manière sensible, surtout pour les grosses sections

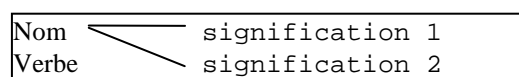


Figure 4 : le cas de (*perte* ; *perdre*) dans le corpus GDP

4.2.5. Nom et verbe ont une signification commune mais le nom a une signification supplémentaire

La paire *échanger/échange* fonctionne sur ce modèle.

élargissement à 2 fois 2 voies des échanges entre rocade est et rocade ouest
échanges de données, d'informations, etc. ; échanger des données, des informations

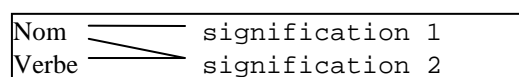


Figure 5 : le cas de (*échange* ; *échanger*) dans le corpus DDE

Ce panorama non exhaustif des alternances possibles entre nominalisations et verbes s'appuie sur l'étude d'exemples où l'interprétation (i.e le classement des contextes argumentaux) est assez facile. Dans beaucoup de cas, l'analyse de ces contextes a été difficile et il a été quasiment impossible de délimiter clairement une signification ou une autre. Le travail à partir de corpus met ainsi en évidence un fonctionnement qui n'est que très rarement évoqué dans les travaux réalisés « par introspection », (Berendonner, 1995, 240) signale pourtant cette difficulté : « Si l'on observe les noms déverbaux en discours, on ne peut manquer d'être frappé par le grand nombre d'occurrences indécidables, en présence desquelles il apparaît non seulement impossible de trancher entre acception concrètes et abstraite mais non pertinent de le faire ». Dans d'autres cas, il s'avère que la distinction classique dynamique vs non-dynamique est insuffisante et devrait être affinée.

4.3. Quelques problèmes

Les problèmes évoqués à la fin du paragraphe précédent : insuffisance de la distinction dynamique vs non-dynamique et indécidabilité quant à la signification vont être détaillés ici à l'aide d'exemples.

4.3.1. Insuffisance de la distinction dynamique vs non-dynamique

Dans de nombreux exemples, la partition classique dynamique/non-dynamique est insuffisante pour rendre compte de la finesse des significations.

Soient les trois énoncés suivants contenant *jalonnement*, extraits du corpus DDE.

Aucun jalonnement n'est mis en place sur le réseau urbain

L'intégration de la signalisation dynamique dans le schéma général de jalonnement des VRU est particulièrement recherchée

Les soumissionnaires devront proposer un emploi du temps détaillé..., ainsi que l'enchaînement et le jalonnement des tâches.

Dans le premier exemple, *jalonnement* correspond à un artefact, il se trouve en position objet d'un verbe demandant un objet concret. Dans le deuxième exemple, *jalonnement* a un sens temporalisé associé à un élément concret (les voies de communication). Enfin, dans le troisième exemple, *jalonnement* a encore un sens temporalisé mais il est associé à un élément abstrait : les tâches. Il se peut d'ailleurs que le deuxième sens de *jalonnement*, sens spécialisé, soit venu contaminé l'usage de *jalonnement* dans le troisième exemple où le mot attendu aurait plutôt été « échelonnement ».

Dans les trois exemples, *jalonnement* a un sens non-dynamique, soit parce qu'il concerne un artefact, soit parce qu'il concerne un état résultatif. La différence entre le premier exemple et les deux autres vient plutôt de ce qu'il reste encore du temps dans les deux derniers, ou plus exactement sans doute de l'aspect. A partir de ces seuls exemples, on constate déjà que la distinction dynamique/non-dynamique n'est pas suffisante et qu'il faudrait l'affiner de la manière suivante :

Dynamique

Non-dynamique, non-temporel,

Non-dynamique, temporel.

Une des prochaines pistes de travail pourrait être la prise en compte de l'aspect dans les nominalisations, ce qui permettrait de mieux rendre compte du fonctionnement en corpus.

4.3.2. Quand le contexte ne permet pas de choisir

Soient les exemples suivants :

Ce dossier d'APS fait suite au Dossier d'études préliminaires actuellement en cours d'approbation.

Les modalités détaillées de financement des équipements, sous réserve de l'approbation de leur cadre...

Responsable de la réalisation des équipements communs, elle soumet pour approbation au Comité de pilotage les différents documents concernant les réalisations.

Dans le premier exemple, *approbation* correspond à un processus qui se déroule, dans le deuxième, *approbation* correspond à un résultat acquis, un état non-dynamique, dans le troisième, il n'est pas possible de décider si *approbation* renvoie à l'un ou l'autre cas. On ne sait pas si les résultats sont soumis au processus d'*approbation* (signification 1) ou s'il est certain qu'ils vont être approuvés (signification 2).

Les cas d'indécidabilité sont fréquents, ce qui est rarement indiqué dans les articles sur le sujet, qui évoquent des distinctions sur des exemples forgés, souvent indiscutables, mais qui laissent malheureusement de côté un grand nombre d'usages réels.

5. Conclusion

Cet article visait à montrer que l'alternance nominalisation/verbe en corpus peut avoir un fonctionnement que la seule introspection ne permet pas de percevoir. Il semble acquis, (deux études l'ont montré) que les nominalisations sont plus utilisées que les formes verbales, dans les corpus spécialisés par rapport à d'autres types de corpus (en tout cas de corpus littéraires et journalistiques). On peut penser que cette sur-utilisation vise à dépersonnaliser l'expression de l'action, d'une part sans doute parce que le discours technique est censé relever d'une connaissance consensuelle et d'autre part parce que la forme nominale est considérée comme une forme plus achevée dans l'abstraction (et donc la scientificité) que la forme verbale. En corpus, l'utilisation de la forme nominale crée très souvent une ambiguïté que le contexte ne permet pas de lever ; ce fonctionnement est rarement perçu par les approches uniquement introspectives. Par ailleurs, dans les cas où le contexte permet de discriminer les significations, on voit apparaître une finesse sémantique qui va bien au-delà des seules valeurs (artefact ou processus) généralement admises et qui nécessite que l'on prenne en compte la notion d'aspect que l'on utilise pour les verbes. Enfin, l'alternance nominalisation/verbe prend en corpus des formes diverses, depuis l'exclusion de l'une ou l'autre forme, jusqu'à l'utilisation de l'une pour une signification et de l'autre pour une autre jusqu'à des polysémies de l'une ou de l'autre qui se recouvrent partiellement. Finalement, on s'aperçoit que l'analyse de corpus oblige à prendre en compte une finesse de fonctionnement sans doute difficile à modéliser mais qui permet de rendre compte au plus juste du fonctionnement des nominalisations et des formes verbales qui leur correspondent. Ainsi, ce type d'étude montre les possibilités qu'ouvre le fait de pouvoir travailler sur des corpus à l'aide d'outils ; cette possibilité correspond certainement à un tournant majeur dans la discipline linguistique.

6. Bibliographie

Bartning, I. (1990) « Les syntagmes binominaux en de - les types interprétatifs, subjectifs et agentifs » ; In actes du Xème congrès des romanistes scandinaves ; *Etudes Romanes de Lund* 45, Lund : Lund University Press. 20-34.

- Bartning, I. (1996) « Les nominalisations déverbales dans les SN complexes en *de* envisagées sous l'angle des traits processifs et résultatif ainsi que de l'opposition abstrait/concret », in N.Flaux, M.Glatigny, D.Samain, (eds) *Les noms abstraits*: Lille : Presses Universitaires du Septentrion, 323-336.
- Beszterda J., Sypnicki J (1986) « Les noms d'action en tant que termes », in *META*, Vol.32, n° 3, p.255-259.
- Berrendonner A. (1995) Redoublement actanciel et nominalisations », in SCOLIA n°5, M.J. Reichler-Béguelin (ed) : Problèmes de sémantique et de relations entre micro et macro-syntaxe. P.215-244.
- Biber D. (1996) : « Investigating language use through corpus-based analyses of association patterns », in *International Journal of Corpus Linguistics*, vol 1(2), 1996, p. 171-197.
- Bosredon, A. (1988) « Un adjectif de trop : L'adjectif de relation ». *L'information Grammaticale* 37, Mars 1988, 60-83.
- Bourigault D., Condamines A. (1999) « Alternance nom/verbe : exploration en corpus spécialisés ». In B.Victorri, J.François (eds) : *Sémantique du lexique verbal*, Actes de l'atelier de Caen, 22-23 janvier 1999. P.41-48.
- Brunet E. (1995) « Un hypertexte statistiques pour grands corpus: HYPERBASE. In lexicomatique et Dictionnaire », *IVes journées scientifiques du réseau lexicologie, Terminologie, Traduction*, Lyon.
- Condamines A. (1998) « Analyse des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus "général" ». In André Clas, Salah Mejri et Taieb Baccouche (eds) : *La Mémoire des mots*, Tunis, Aupelf-UREF, SERVICED, Coll. Actualités scientifiques. 351-367.
- Condamines A. (1999) « Analyse des structures « Nominalisation + de N », « Nominalisation + N », « Nominalisation + Adjectif relationnel » dans un corpus spécialisé ». In M.Plénat et M.Aurnague (eds) *L'emprise du sens, structures linguistiques et interprétations*, Amsterdam, Atlanta : Rodopi ; p. 61-82.
- Defranc, B. ; Willems, D. (1996) « De l'abstrait au concret, une réflexion sur la polysémie des noms déverbaux ». in N.Flaux, M.Glatigny, D.Samain (eds), *Les noms abstraits* : Lille : Presses Universitaires du Septentrion, 221-230.
- Fabre, C. (1996) « Interprétation automatique des séquences binominales en anglais et en français. Applications à la recherche d'informations ». Thèse de l'Université de Rennes 1, 1996.
- Fabre C. (1998) : « Repérage de variantes dérivationnelles de termes ». Carnets de grammaire, n°3, Equipe de Recherche en Syntaxe et Sémantique.
- Habert, B. ; Nazarenko ; A., Salem, A. (1997) « Les linguistiques de corpus ». Paris : Armand Colin.
- Mélis-Puchulu, A. (1991) « Les adjectifs dénominaux : des adjectifs de "relation" » ; *Lexique* n°10, 1991, 33-60.
- Noailly M. (1990) « Le substantif épithète », Paris : PUF, 1990.
- Rémi-Giraud Sylviane (1996) « Pour une approche notionnelle de la nominalisation », in *Les noms abstraits, histoire et théorie*, Presses Universitaires du Septentrion, 1996; p.105-116.
- Samvelian, P. (1995) « Les nominalisations en français : arguments sémantiques et actants syntaxiques » ; Thèse de l'Université Paris 7.
- Rastier F. (1995) « Le terme entre ontologie et linguistique », in *La Banque des mots*, numéro spécial 7, actes des premières rencontres "Terminologie et Intelligence Artificielle", 1995. pp.35-64.
- Wuster E. , 1981, " L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses ", dans Rondeau G. et Felber H. (eds), *Textes choisis de terminologie*, GIRSTEM, Université Laval à Québec, pp. 57-114.